What is CDNOT and what is it we want to achieve ?

Sébastien Denvil Institut Pierre Simon Laplace, IPSL, France

What is the CDNOT ?

A CMIP Data Node Operations Team has been appointed by the WGCM Infrastructure Panel (WIP) and include representatives from groups responsible for CMIP data nodes.

The CDNOT scope covers functions and resource issues (hardware, networks, people) related to: installing, configuring and operating all the nodes and node services in the CMIP data federation and the policies and processes involved in both managing data on a data node (including acquisition, quality assurance, citation, versioning, and publishing) and providing access services.

What is the CDNOT ?

Providing formal input to the WIP with regard to data node operations, resource issues, and other concerns related to meeting WIP requirements.

Providing formal input to any software providers as to software requirements and constraints arising from the operational environment/experience.

Maintaining documentation outlining the required resources, configuration and mandatory access services on a data node.

Agreeing how to configure federation-level services and service nodes to support WIP requirements.

Where necessary, establishing a timetable and methodology for federation-wide software migrations, upgrades or configuration changes.

What is the CDNOT ?

Drafting and maintaining guidelines describing how the CMIP data management plans should be implemented on data nodes (e.g. to cover how identifiers should be obtained and used, metadata requirements, quality control and the addition and removal of data with versioning).

Monitoring for compliance the implementation of the CMIP data management plans across the federation, and providing feedback to the WIP.

FAIR DATA

- FAIR : Findable Accessible Interoperable Reusable
- Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process. Unfortunately, the existing digital ecosystem surrounding scholarly data publication prevents us from extracting maximum benefit from our research investments
- Wilkinson et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016 Mar 15;3:160018. doi: 10.1038/sdata.2016.18,

Feeling the heat ?

All dates in red are official dates from IPCC plenary in Nairobi, 2016-04, and IPCC XC meeting 2016-05-19.

2022-09: AR6 Synthesis Report

2021-02: WG1 Report Approved

2020-12: Final WG1 Draft goes to inter-governmental review

2020-06: 4th Lead Author meeting

2020-02: Post 3rd LA meeting, second-order draft sent out for expert review. Any citations here will have to have been submitted for peer review by this date. First-order draft can use pre-citation material.

2018-12: IPSL runs complete.

Early 2018 : IPSL data starts to flow in.

Earlier special reports (1.5C, cryosphere, land) not based on CMIP6.

What is expected ?

The WIP and the IPCC would like to be able to anticipate our data provisioning. It will be valuable for all if we can give on overview on where we are and what is planned with respect to data production in our centers. How much simulation currently running, expected date of completion, expected date of standardisation and expected date of publication on ESGF.

We decided we will try our best being informative but will be very cautious not to trigger unnecessary expections.

High level readiness

Institution	Model ready	Spin up started	Deck simulation started	Post processing workflow ready	Data ready for ESGF	Data on ESGF
IPSL-CM6-LR	Yes	Yes	Almost	Yes	Early 2018	Early 2018
CCCma	Testing	Testing	No	Almost	No	No
CCCR-IITM	Almost	Yes	NO	Almost	NO	NO
DOE (ACME)	Almost	NO	NO (September 2017)	Almost	NO	NO
NASA GISS	Yes (GISS-E2.1, late 2017)	Yes	Almost (Dec 2017)	Yes	NO (Feb/Mar 2018)	NO (Feb/Mar 2018)
CNRM-CM6-1	Yes	Yes	almost	yes	yes (test data) / almost (deck data)	waiting for stack release
CNRM-ESM2-1	Almost	Yes	NO	yes	NO	NO
CNRM-CM6-1-HR	NO	NO	NO	Almost	NO	NO
CNRM-ESM2-1-HR	NO	NO	NO	Almost	NO	NO
MOHC (HadGEM3-GC31-LL)	Yes	Yes	Yes	Almost	NO	NO
MOHC (HadGEM3-GC31-MM)	Yes	Yes	Yes	Almost	NO	NO
MOHC (UKESM1-0-LL)	Almost	Yes	NO (expected Dec 2017)	Almost	NO	NO
NOAA GFDL (CM4)	Almost	NO	NO	Almost	NO	NO
NOAA GFDL (ESM4)	NO	NO	NO	Almost	NO	NO
MPIESM1 (LR)	Almost	Yes	NO	NO	NO	NO
MPIESM1 (HR)	Almost	Yes	Almost (Jul 2017)	NO	NO	NO
MPIESM2	NO	NO	NO	NO	NO	NO
DLR	Yes	Almost (August 2017)	Almost (August 2017)	Almost	NO	NO
NorESM	Almost	NO	NO	Almost	NO	NO
AWI-CM	Yes (since August 2017)	Yes (since August 2017)	NO (December 2017)	NO (August 2017)	NO (March 2018)	NO (end 2018)
NCAR-CESM	Almost	Yes	NO	Almost	NO	NO
CMCC-ESM2	Almost	NO (fall 2017)	NO (fall/winter 2017)	Almost	NO	NO
CMCC-CM2-HR4	Yes	Yes	NO	Almost	NO	NO
CMCC-CM2-HR5	NO	NO	NO	NO	NO	NO
CMCC-CM2-SR5	Yes	Yes	NO (fall/winter 2017)	Almost	NO	NO
EC-EARTH	NO	NO (late 2017)	NO (late 2017)	NO	NO	NO
CAS FGOALS	Almost	NO	NO	Almost	NO	NO

The Workflow

Documentation → Production → Standardisation

Standardisation → Quality control → Data Reference Syntax → Versioning

Versioning → Mapfiles → Metadata catalogs → ESGF indexing

Green light is green light

ESGF test suite goals:

- Test ESGF stack automatically and produce a nice html report
- Modular testing: ready to use sets of tests or create new sets of tests
- Bring a test framework for the ESGF community
- A step further to the Continuous Integration of the ESGF-docker release

Properties:

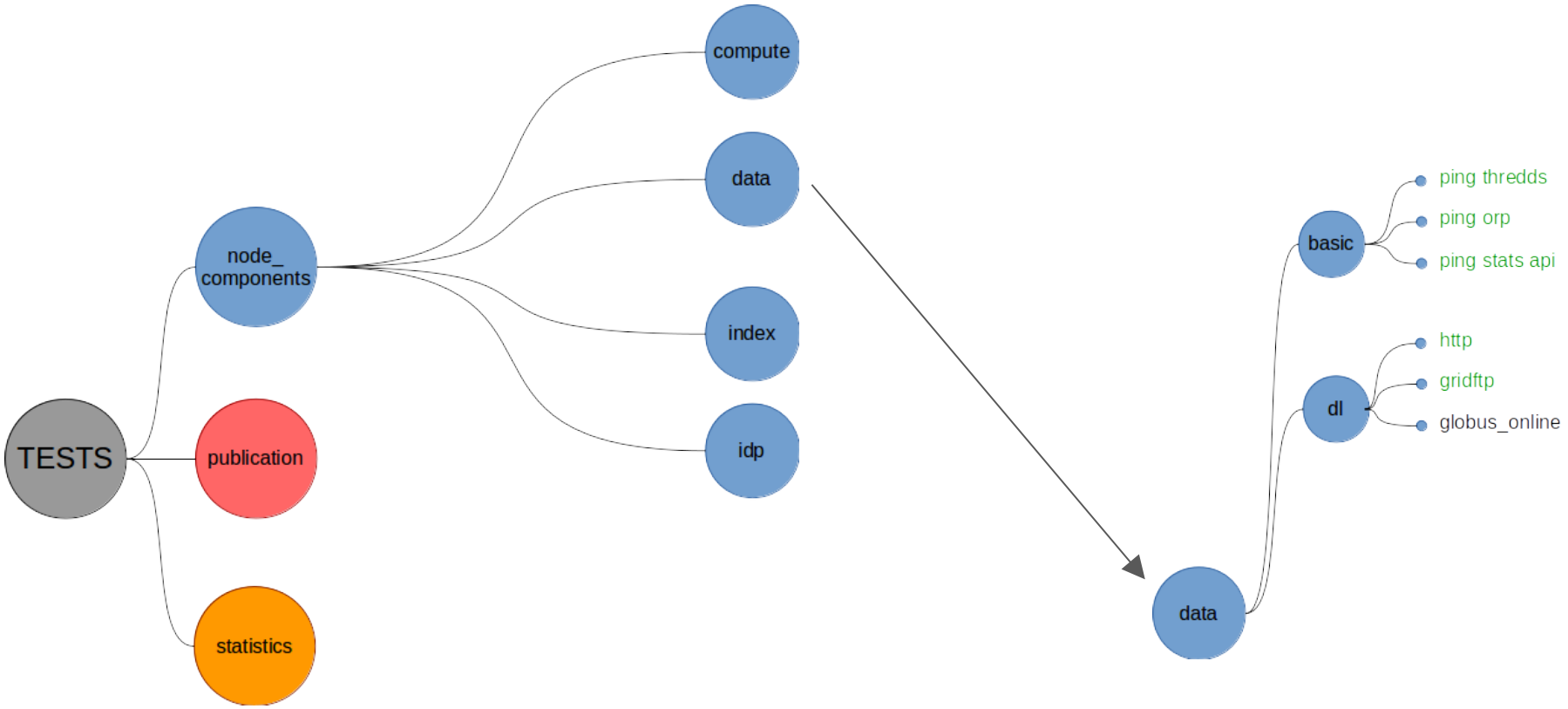
- Implementation use the well known nose (Python)
- Standalone and integrable into Hudson (CI)
- Tests are organized like a mind map

git clone https://github.com/ESGF/esgf-test-suite.git
https://github.com/ESGF/esgf-test-suite/blob/master/README.md

We have to capitalyze tests

Mind map of tests

- non-terminal nodes are sets of tests
- terminal nodes are test cases
- set operators are availables
- test cases can be run independently



CMIP6 monitoring

Data Node	Nov 30			Dec 1			Dec 2			Dec 3			Dec 4			Dec 5		
	HTTPS	GridFTP	Globus	HTTPS	GridFTP	Globus	HTTPS	GridFTP	Globus	HTTPS	GridFTP	Globus	HTTPS	GridFTP	Globus	HTTPS	GridFTP	Globus
aims3.llnl.gov																		
cordexesg.dmi.dk																		
esg-cccr.tropmet.res.in																		
esg-dn1.nsc.liu.se																		
esg.cnrm-game-meteo.fr																		
esgf-data.jpl.nasa.gov																		
esgf-data.ucar.edu																		
esgf-data1.ceda.ac.uk																		
esgf-data1.diasjp.net																		
esgf.anl.gov																		
esgf.ichec.ie																		
esgf.nccs.nasa.gov																		
esgf.nci.org.au																		
esgf1.dkrz.de																		
noresg.norstore.no																		
prodn.idris.fr																		
vesg.ipsl.polytechnique.fr																		
vesg.ipsl.upmc.fr																		

CMIP6 monitoring

Certificate Expiration Date

Data Node	HTTPS	GridFTP
aims3.llnl.gov	2018-07-25 23:59:59	2018-09-26 09:39:08
cordexesg.dmi.dk	2018-02-13 09:53:28	
esg-cccr.tropmet.res.in	2018-07-12 16:03:38	
esg-dn1.nsc.liu.se	2019-01-30 12:00:00	2017-12-30 12:37:04
esg.cnrm-game-meteo.fr	2020-03-27 12:00:00	
esgdata.gfdl.noaa.gov	2018-03-16 10:19:52	2018-03-16 10:19:52
esgf-data.jpl.nasa.gov	2018-07-19 12:00:00	
esgf-data.ucar.edu	2018-12-03 05:06:54	2018-12-03 05:06:54
esgf-data1.ceda.ac.uk	2018-07-01 11:11:08	
esgf-data1.diasjp.net	2019-08-13 14:24:46	
esgf-data2.ceda.ac.uk	2018-07-01 11:12:16	
esgf.anl.gov	2018-06-27 22:00:54	2019-05-28 01:08:21
esgf.ichec.ie	2018-02-18 11:07:55	
esgf.nccs.nasa.gov	2019-05-04 12:00:00	
esgf.nci.org.au	2019-03-11 04:41:32	
esgf1.dkrz.de	2019-04-25 13:52:11	2018-10-13 10:03:18
noresg.norstore.no	2019-06-21 12:00:00	
prodn.idris.fr	2018-06-25 23:59:59	
vesg.ipsl.polytechnique.fr	2020-03-31 12:00:00	2018-08-22 13:20:24
vesg.ipsl.upmc.fr	2018-03-07 12:00:00	2018-01-04 15:56:30

IdP/Index Node	HTTPS	MyProxy
esg-dn1.nsc.liu.se	2019-01-30 12:00:00	2017-12-30 12:37:04
esgdata.gfdl.noaa.gov	2018-03-16 10:19:52	2018-03-16 10:19:52
esgf-data.dkrz.de	2019-04-05 09:32:11	2018-10-13 10:05:56
esgf-node.ipsl.upmc.fr	2018-03-07 12:00:00	2018-01-04 15:14:30
esgf-node.jpl.nasa.gov	2018-07-19 12:00:00	2019-12-14 20:55:06
esgf-node.llnl.gov	2018-03-14 23:59:59	2019-04-19 22:36:24
esgf.esrl.noaa.gov	2020-08-31 12:00:00	
esgf.nccs.nasa.gov	2019-05-04 12:00:00	
esgf.nci.org.au	2019-03-11 04:41:32	2018-02-13 09:34:20

Legend
Expired or another error
Expires within 2 weeks
Expiration time

Take home messages

Be more specific on which Tiers 1 support which Tiers 2

Are 2 weeks to upgrade reasonable for Tier 2 nodes? Attendees thought yes. We think it goes with the ability to remove a node from the federation if necessary.

CDNOT will have to establish a step by step procedure before removing a node. Remove a node from the federation can be temporary if the issue is fixed later on.

Take home messages

2.5.16 has been released and fixed afterwards. Not easy for node manager outside the core circle of developer to find a way within the ESGF release cycle. We decide to communicate clearly to CDNOT when a release is available and rock solid.

ESGF release cycle versioning document needs to be clear outside this room also.

Storage dedicated replication

PCMDI 5 years plan. Up to 20 PB spinning by late 2020.

ANU Hope to contribute up to 10 PB. (unsure spinning disk or tape?)

DKRZ 4PB spinning disk, > 20 PB offline (tape, but ESGF publishable and accessible via ESGF data node)

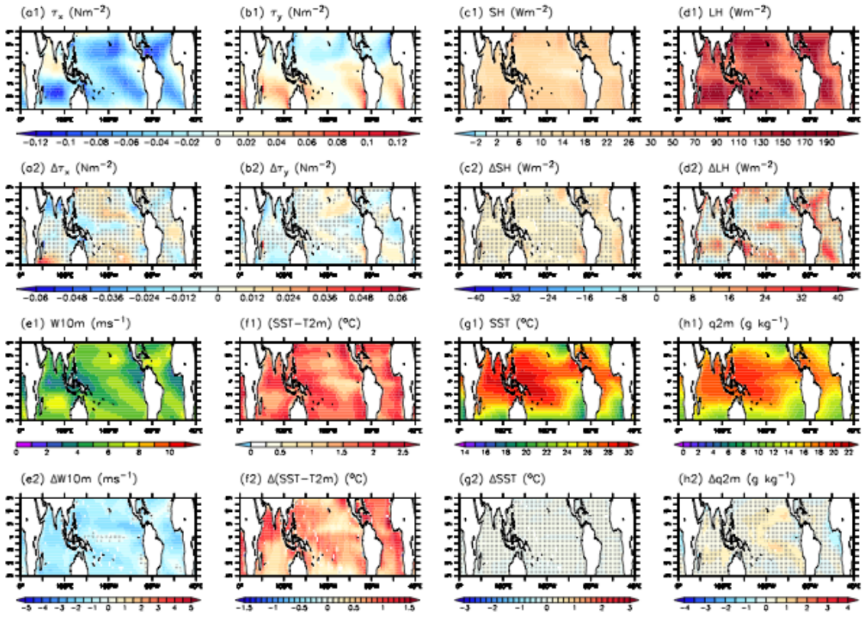
IPSL 4 PB spinning disk.

CEDA 2 PB spinning disk. > 2 PB offline (tape, but ESGF publishable and accessible via ESGF data node)

BNU up to 3PB

FGOALS up to 3PB.

Questions ?


(a1) τ_x (Nm^{-2})
(b1) τ_y (Nm^{-2})
(c1) SH (Wm^{-2})
(d1) LH (Wm^{-2})
(a2) $\Delta\tau_x$ (Nm^{-2})
(b2) $\Delta\tau_y$ (Nm^{-2})
(c2) ΔSH (Wm^{-2})
(d2) ΔLH (Wm^{-2})
(e1) W10m (ms^{-1})
(f1) (SST−T2m) (°C)
(g1) SST (°C)
(h1) q2m (g kg^{-1})
(e2) ΔW10m (ms^{-1})
(f2) Δ(SST−T2m) (°C)
(g2) ΔSST (°C)
(h2) Δq2m (g kg^{-1})